Journal of Hydrology 612 (2022) 128262

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Groundwater level prediction with meteorologically sensitive Gated Recurrent Unit (GRU) neural networks

Amin Gharehbaghi [a], Redvan Ghasemlounia [b], Farshad Ahmadi [c], Mohammad Albaji [d,*]

[a] Faculty of Engineering, Dept. of Civil Engineering, Hasan Kalyoncu University, Şahinbey, Gaziantep Postal Code: 27110, Turkey
[b] Faculty of Engineering, Dept. of Civil Engineering, Istanbul Gedik Univ, Istanbul Postal Code: 34876, Turkey
[c] Department of Hydrology & Water Resources Engineering, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran
[d] Department of Irrigation and Drainage, Faculty of Water and Environmental Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran

ARTICLE INFO

This manuscript was handled by Corrado Corradini, Editor-in-Chief, with the assistance of Gokmen Tayfur, Associate Editor

Keywords:
Regional mean monthly groundwater level
GRU neural network
Shannon entropy method
Cosine amplitude sensitivity analysis

ABSTRACT

Precise estimation of groundwater level (GWL) fluctuations has a substantial effect on water resources management. In the present study, to forecast the regional mean monthly time series groundwater level (*GWL*) with a range of 4.82 (m) in Urmia plain, three different layer structures of Gated Recurrent Unit (GRU) deep learning-based neural network models via the module of sequence-to-sequence regression are designed. In this sense, 180-time series datasets of regional mean monthly meteorological, hydrological, and observed water table depths of 42 different monitoring piezometers during the period of Oct 2002–Sep 2017 are employed as the input variables. By using Shannon entropy method, the most influential parameters on *GWL* are determined as regional mean monthly air temperature (T_{am}), precipitation (P_m), total (sum) water diversion discharge (W_{dm}) of four main rivers. Nevertheless, Cosine amplitude sensitivity analysis confirmed T_{am} as a dominant factor. For preventing overfitting problem, an algorithm tuning technique via different kinds of hyperparameters is operated. In this respect, several scenarios are implemented and the optimal hyperparameters are accomplished via the trial-and-error process. As stated by the performance evaluation metrics, Model Grading process, and Total Learnable Parameters (*TLP*) value, the innovative and unique suggested model (3), entitled GRU2+, (Double-GRU model coupled with Addition layer (+)) with seven layers is carefully chosen as the best model. The unique suggested model (3) in the optimal hyperparameters, resulted in an R^2 of 0.91, a total grade (*TG*) of 7.76, an *RMSE* of 0.094 (m), and a running time of 47 (s). Thus, the model (3) can be certainly employed as an effective model to forecast *GWL* in different agricultural areas.

1. Introduction

Groundwater considers a noteworthy freshwater supply all over the world. Howbeit, this natural source has confronted with diverse problems such as anthropogenetic processes, climatical change consequences, and naturalistic events that menace its availability (Houéménou et al., 2020). Overconsuming groundwater resources by industry and farming sections has brought about extreme water famine and long-lasting hydrological and ecological disasters (Alabjah et al., 2018).

A groundwater level (*GWL*) is a consequence of diverse time-limited and autonomous procedures with unnatural and natural basics. Accurate prediction of *GWL* fluctuations trend is regarded as a fundamental process to recuperate the groundwater resources and functional supervision by monitoring the waterway. But, due to uncertainties in the hydrogeological and meteorological factors, nonlinear interactions, and some human activities, defining a mathematical method to determine *GWL* over a long period in the agrarian districts is contemplated as an arduous undertaking and also a hot subject for many scholars (Raghavendra and Deka, 2016).

Routinely, a hydrogeological process can be scrutinized by three approaches, including physical, statistical, and artificial intelligence (*AI*) methods. Lately, by perceptible progressions in computer knowledge, some sophisticated mathematical methods have been presented concerning environmental and groundwater hydrology (Gharehbaghi, 2016, 2017, 2022; Dalkiliç and Gharehbaghi 2021; Chen et al., 2017).

* Corresponding author.
E-mail addresses: amin.gharehbaghi@hku.edu.tr (A. Gharehbaghi), redvan.ghasemlounia@gedik.edu.tr (R. Ghasemlounia), f.ahmadi@scu.ac.ir (F. Ahmadi), m.albaji@scu.ac.ir (M. Albaji).

https://doi.org/10.1016/j.jhydrol.2022.128262
Received 11 May 2022; Received in revised form 12 July 2022; Accepted 21 July 2022
Available online 28 July 2022

Numerical methods were derived from the partial differential equations, yet they have some own restrictions and hitches such as the buildup of physical data, especially for the long-term *GWL* forecasting (Daliakopoulos et al., 2005).

To date, to lessen the difficulties and limitations pertinent to numerical methods, some standalone Data-Driven methods (DDMs), Machine Learning-based models (MLMs) have been developed for forecasting *GWL* (Shiri et al., 2013; Sahoo and Jha, 2013; Mohanty et al., 2013; Meshgi et al., 2014; Mirzavand et al., 2015; Chang et al., 2016). In spite of these methods were characterized as fairly flexible and systematic schemes, some shortcomings have been reported such as overfitting problem and extreme uncertainty (Yoon et al., 2011). As well, since the non-stationary time series *GWL*, the conventional artificial neural networks (ANNs) like Feed-Forward Neural Network (FFNN) models were encountered serious difficulties such as weights adjustment and lack of capabilities for remembering prior information (Coulibaly et al., 2001; Feng et al., 2008).

To exclude the all-above-mentioned problems, novel structures of big data analytics by a specific kind of Deep Neural Network (DNN) based models were developed to address precisely the long-term randomization and non-stationery nature of *GWL*. They have been renewed as the innovative generation of ANNs in numerous scientific disciplines in recent years. Generally, DNN models have interior self-looped cells which cause to recall the prior info and make it efficient for performing time series-based tasks (Graves, 2012). They represent a substantial performance in the capability of neural networks (NNs) for mechanically engineering problem-related aspects and learning extremely intricate data distributions in the fields of water resources and hydrology (Shen, 2018). The foremost advantage of modelling by DNNs is that physical relationships and variables required for defining the physical environment do not necessitate to be clearly described. Concerning, they evaluate the relationship between model input/output by a reiterative aspect of the learning process (Solomatine and Obstfeld, 2008).

As this study focuses on a time series forecasting task and the behavior modelling also includes temporal dependencies, consequently, Gated Recurrent Unit (GRU) neural network, DNNs-based, could be contemplated as an opposite option (Rumelhart et al., 1986). GRU was recommended by Cho et al. (2014) and evolved as a state-of-the-art architecture of recurrent neural networks (RNNs). It has a comparable chain-form building with conventional RNNs, yet the interior functions in GRU cells are further elaborated by supplementary cell states in which information can be kept. It can successfully forestall the inherent gradient disappearance difficulty in the conventional neural network under the historical data by using relevant internal gates (Pan et al. 2020). Practically, GRU as an extemporized form of RNNs is proficient in learning long-term dependencies and sequence data modelling on condition that the data volume is not too big (Cho et al. 2014). In fact, GRU ponders as a modification type of Long Short-Term Memory (LSTM) as the most common version of DL-based neural network. Albeit both models perform in the same way, GRU deliberates as a stand-in neural network because it is calculably cheaper, a quicker learning curve, more condensed construction, and less learnable parameters (Zhang et al., 2018; Park et al., 2019).

In the present study, different layer structures of GRU model are designed that the new suggested robust DL-based structure (3), considers an innovative coupled version of GRU neural network. It is willfully developed to increase the estimation accuracy of the time series *GWL*. Thus, we do not confine our analysis only to the traditional network structure of GRU model.

Lately, LSTM neural network model has been frequently utilized in the field of hydrological science by majority of scholars to forecast lake water level (Liang et al., 2018), short-term daily reference evapotranspiration (Yin et al., 2020), and water table depth in different areas (Zhang et al., 2018; Bowes et al., 2019; Vu et al. 2020; Jeong et al. 2020). Jeong and Park (2019) employed different DNN methods to forecast *GWL*. The results confirmed that LSTM and NARX (Nonlinear Auto-Regressive Exogenous) neural network models were more precise than ARX and GRU models. Chen et al. (2020) developed a substitute method of groundwater modeling via GRU neural network model to progress the effectiveness of parameter auto-calibration and global sensitivity analysis. The results attested that GRU-based adoptive model showed a high capability to handle dilemmas with multiple time series output. Gao et al. (2020) forecasted short-term runoff with GRU and LSTM neural networks without using time step optimization in sample generation. They reported that GRU model yielded in equivalent results as well as LSTM model, however, GRU model was selected as a favored model, as it needs less time for model training. Ghasemlounia et al. (2021) forecasted *GWL* fluctuations by operating Bi-directional Long Short-Term Memory (BiLSTM) deep neural network without employing meteorological and hydrological variables. They reported the proposed model (simple Double-BiLSTM model combined by Multiplication layer ($\times$)) with the module of sequence-to-one as the superior model. This model in the piezometer 4 with a range of 4.49 (m) resulted in an R^2 of 0.89 and an *RMSE* of 0.17 (m), yet in the similar physical features, the simple Double-BiLSTM model resulted in an R^2 of 0.77 and an *RMSE* of 0.25 (m). Wunsch et al. (2021) compared the ability of different artificial neural networks, including LSTM, convolutional neural networks (CNNs), and NARX neural network in forecasting 17 groundwater wells within Upper Rhine Graben region. They equated the performance of developed models for predicting on 4-year period by using both sequence-to-one (seq2val) and sequence-to-sequence (seq2seq) modules via both GWL_{t-1} and existing meteorological as the input parameters. The results confirmed that in the lack of training information, NARX model was generally outclassed in both seq2val and seq2seq forecasts module compared to the advanced DL techniques.

Although the function of LSTM is a growing investigation domain, the performance of different layer structures of GRU neural network model for predicting more complex natural phenomena is scarce, yet. To the best of the authors' knowledge, despite various examinations in the prediction of *GWL* using the traditional structure of DL-based models, no investigation has been executed in the literature on probing the function of exclusive suggested model. The novelty of this research is developing an advanced GRU-based neural network structure with the module of sequence-to-sequence regression for the prediction of intricate natural phenomena such as time series *GWL* fluctuations.

The scope and main aims of the current study are as follows: 1) to develop different layer structures of GRU based neural network models for precise prediction of the regional mean monthly *GWL* using observed long-period water table depth of several piezometers, hydrological, and meteorological datasets with time series characteristics in Urmia plain, West Azerbaijan, Iran; 2) to specify the optimal amount of intended hyperparameters such as kind of state activation function (*SAF*), number of neurons (*NoN*), network architectural design *(NAD)*, and dropout rate (*P-rate*) for well-configuration of designed models to lessen the impact of overfitting problem; and 3) to evaluate and analogize outcomes of modelling to single out the best-designed model by statistical metrics.

2. Materials and methods

2.1. Study area and data acquisition

Urmia Plain is placed on the western side of Urmia Lake, in Urmia City, West Azerbaijan Province, Northwest of Iran, among 44°55'-45°18'E longitude and 37°20'-37°49'N latitude coordinates with 1310 (m) above sea level. It has very fertile land, a cold, semi-arid climate, and an area of about 98,100 ha. Nazloochaei, Rozehchaei, Shahrchay, and Barandozchay are the main rivers and chief surface water stores to irrigate this region that after crossing the western highlands flow into Urmia Lake. The total mean yearly water diversion inflow discharge by main rivers is approximately 830E + 6 (m^3). Fig. 1 describes the geographical location map of the study area.

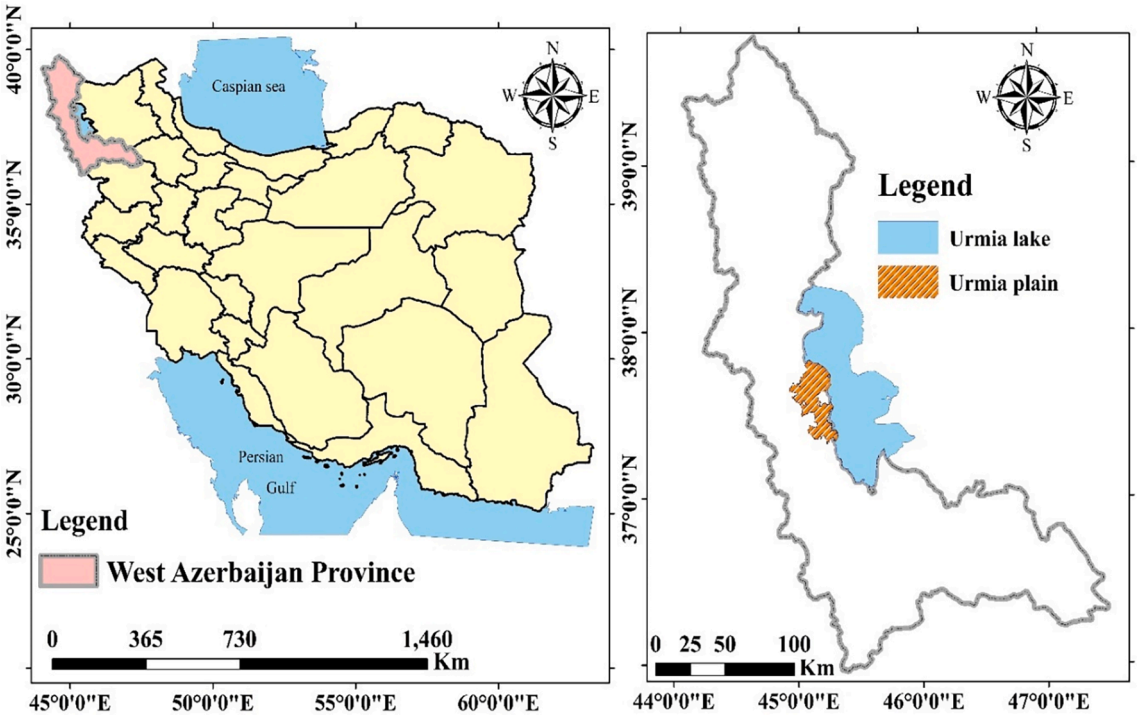


Fig. 1. Geographical location map of the study area.

In summer season the amount of real evapotranspiration (ET_a) is high, so groundwater and surface stores are used as supplementary supplies along with local precipitations. The mean annual *ETPot* is 800–1400 (mm). Peak rainfall is mostly concentrated in the months of January to May. Mean annual precipitation and temperature were reported almost 304 (mm) and 13.5 (°C), respectively (EARWO, 2020). Therefore, the area is exposed to water drought and consequently, groundwater has a significant influence. Urmia plain aquifer is unconfined wherein the stream direction of groundwater is from the west on the way to Urmia Lake.

In the current research, to estimate the time series *GWL* in the study area, time series datasets of 180 monthly meteorological, observed water table depths of 42 different piezometers, and hydrological datasets during the period of Oct 2002–Sep 2017 are used. The meteorological variables are regional average monthly air temperature (T_{am}), precipitation (P_m), and evaporation (ET_m) collected from different meteorological stations and the used hydrological variable is the regional total (sum) mean monthly water diversion discharge (W_{dm}) of the main rivers collected from different hydrometry stations positioned.

Fig. 2 illustrates the site of the observation piezometers, meteorological, and hydrometry stations in the study region together with Urmia Lake, basin, and main rivers. As can be seen from Fig. 2, entirely 42 monitoring piezometers are applied so as to cover all study regions with randomly observed water table depth.

Time series plots of regional mean monthly *GWL*, meteorological, and hydrological variables between Oct 2002–Sep 2017 are presented in Fig. 3 (A-E). According to Fig. 3, the variables used have a time dimension, so time in the monthly dimension should be utilized as an

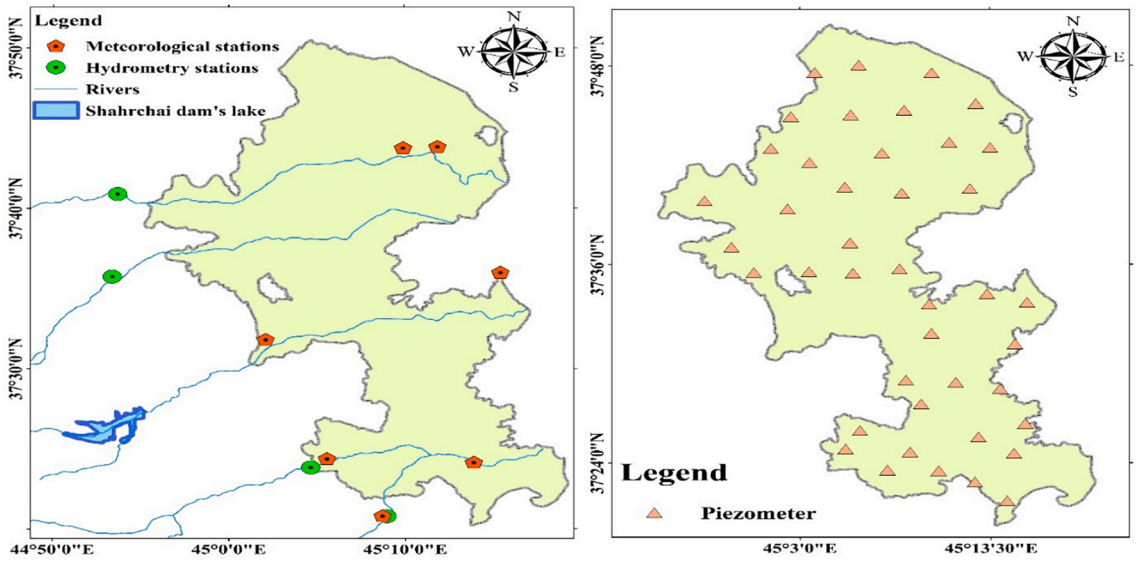


Fig. 2. Place of the piezometers and meteorological and hydrometry stations in Urmia plain, West Azerbaijan Province, Iran.

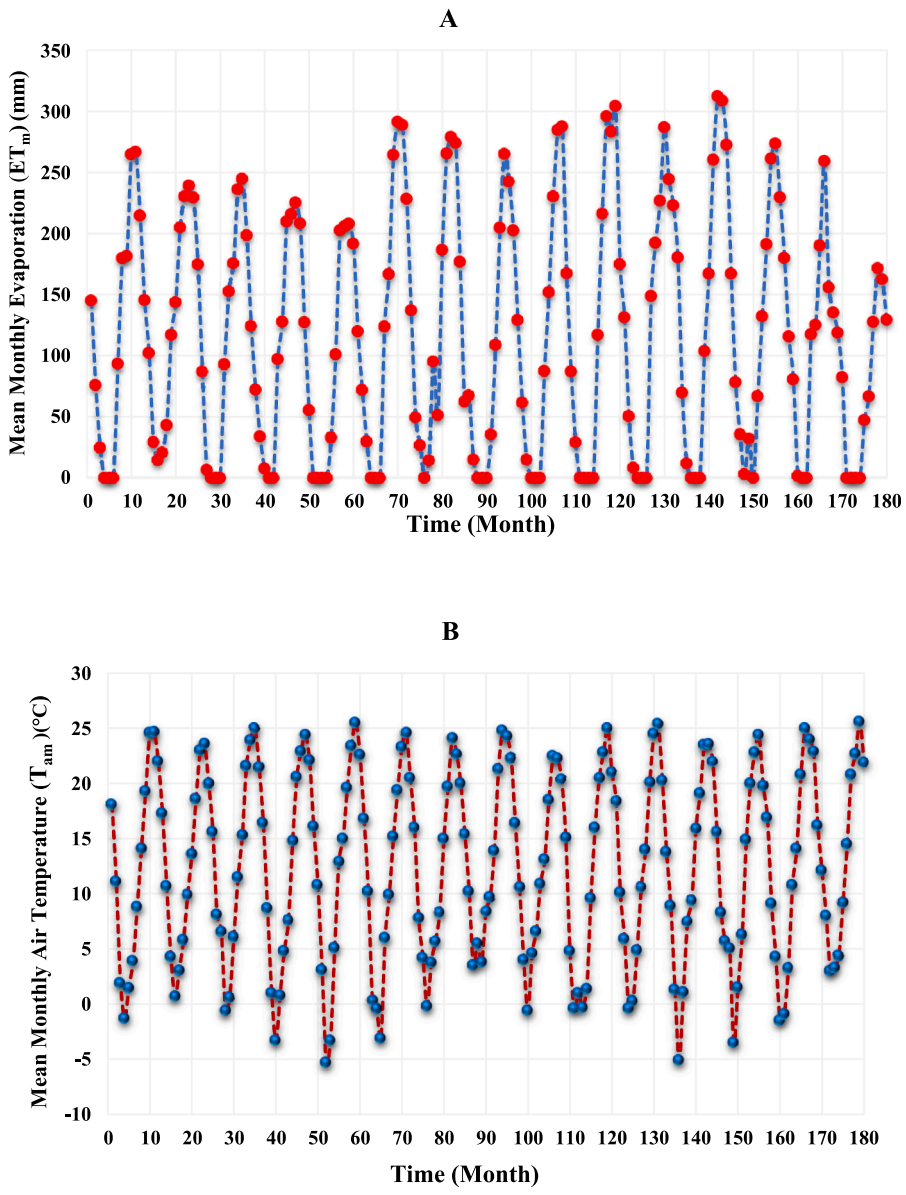


Fig. 3. Time series plots of mean monthly used variables between Oct 2002–Sep 2017 in the study district (Urmia plain) including: (A) Evaporation, (B) Temperature, (C) Water Diversion, (D) Precipitation, and (E) mean monthly *GWL*.

independent factor.

Table 1 presents some descriptive statistics of the variables used wherein *CV* and *STDV* signify the coefficient of variation and standard deviation, respectively. According to Table 1, the range of *GWL* between Oct 2002–Sep 2017 is 4.82 m.

2.2. Data preprocessing and sensitivity analysis

The collected datasets are appraised to determine the most operative variables on *GWL* between Oct 2002–Sep 2017 in the study area. Since the performance of any modelling technique in the accurate forecasting of target primarily relies on an appropriate selection of predictors, inapt choosing could undesirably affect the performance of methods. Regarding, in the current investigation, via a weighting method, the most important input variables are determined. Finding a suitable weight for each variable has been regarded as one of the chief principles in Multi-Attribute Decision Making (MADM) problems. Concerning, Shannon entropy (shortage of certainty of some measures) approach which was introduced by Shannon (1948), is one of numerous methods for obtaining weights conferred in the literature. Because the continuous distribution of a random variable varies from skewed to normal probability, the entropy value increases. Consequently, a small value of entropy signifies a futile random variable and more irregularity in a system (Shannon 1948). Shannon entropy can be computed by following steps (Shannon 1948):

Step 1: Normalize the decision matrix:

$$f_{ij} = \frac{x_{ij}}{\sum_{j=1}^{m} x_{ij}}, (j = 1, \cdots, m, i = 1, \cdots, n) \quad (1)$$

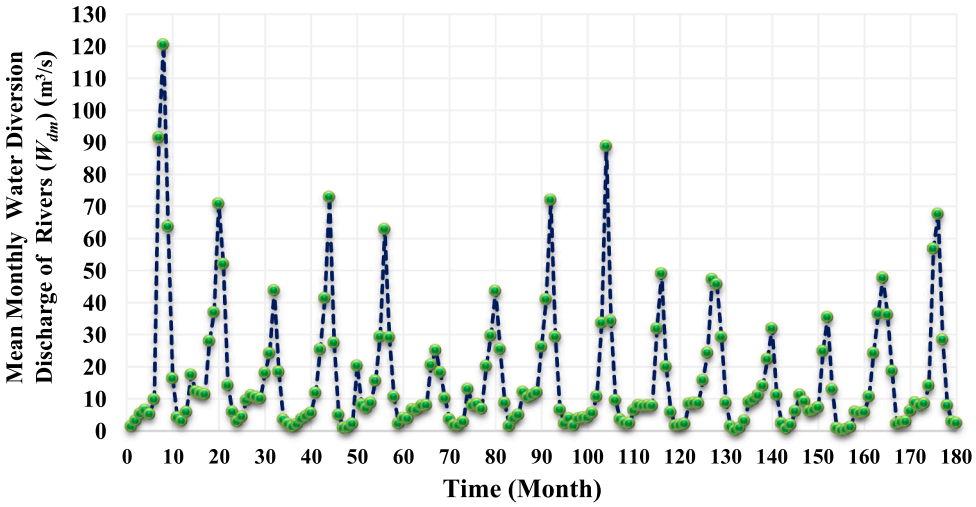
Step 2: Calculate entropy:

$$E_i = -k \sum_{j=1}^{m} f_{ij} Lnf_{ij} \quad (2)$$

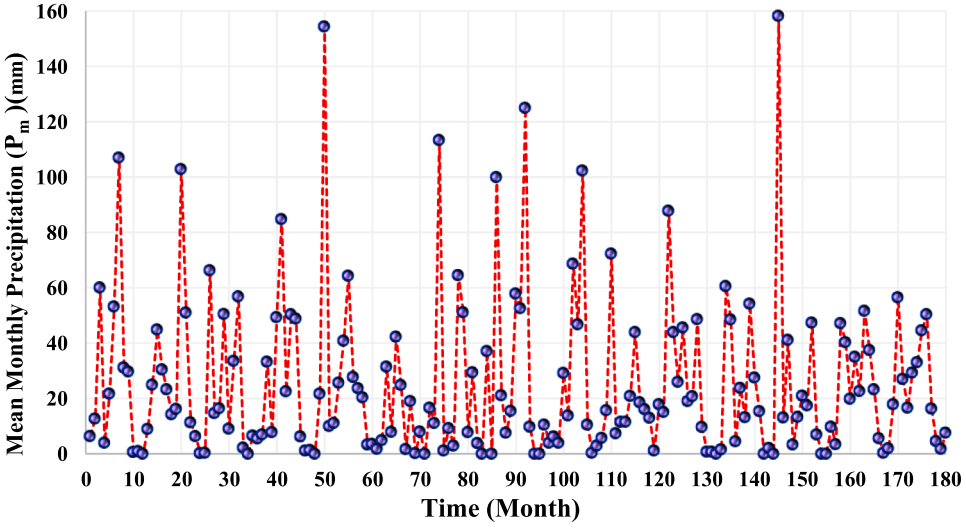
$$k = \frac{1}{Lnn} \quad (3)$$

where x_{ij} is the rating of meteorological or hydrological station i

C



D



E

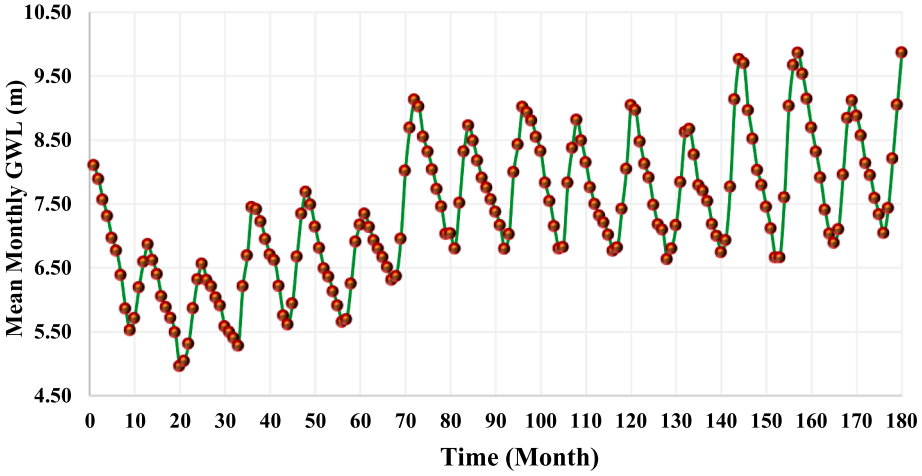


Fig. 3. (*continued*).

Table 1
Statistical parameters of used time series variables between Oct 2002–Sep 2017 in the study area.

Variables	Data period	Min	Max	Mean	STDV	Skewness	CV
$ET_{m(mm)}$	Whole period	0	312.5	116.66	98.44	0.29	0.84
T_{am} (°C)	Whole period	−5.3	25.6	12.37	8.7	−0.13	0.7
P_m (mm)	Whole period	0	158.34	25.44	28.86	2.01	1.13
W_{dm} (m^3/s)	Whole period	0.25	120.53	16.07	19.34	2.38	1.2
GWL (m)	Whole period	4.96	9.78	7.37	1.08	0.12	0.15

concerning variable j, m is the number of variables, f_{ij} is the normalized x_{ij}, n is the number of stations, k is the entropy constant, and E_i is the amount of entropy for variable i.

Step 3: Determine uncertainty:

$$d_i = 1 - E_i \tag{4}$$

where d_i show the degree of deviation or uncertainty of the data for variable i.

Step 4: Determine the significance of variable i:

$$\widehat{W}_j = \frac{d_i}{\sum_{i=1}^{m} d_i} \tag{5}$$

where W_j indicates the weight vector of variable j.

Table 2 presents the computed Shannon entropy weight value for used time series meteorological and hydrological variables with respect to *GWL*. The computed weighting coefficients show the comparative significance and heterogeneousness of each selected variable for predicting the time series *GWL*. On the basis of Table 2, the effect of ET_m on *GWL* is discardable because of its minimal total weight value and consequently least impact. Thus, P_m, T_{am}, and W_{dm} elite as the ultimate effective input variables; Which, among these variables, T_{am} owing to the maximum total weight value has a high degree of status and contributes strongly in forecasting *GWL* as the methods target.

By discounting the effects of ET_m for small Shannon entropy weight value, *GWL* in final functional form would be as follow,

$$GWL = f(Tam, Pm, Wdm) \tag{6}$$

Equation (6) is a time series of multi-input and one-output variables and considers capable of modelling by DNNs to forecast *GWL*.

A sensitivity analysis for variables used in Equation (6) is adopted to explain the impression of each input variable on the model target by varying each input variable in a fixed rate and preserving the other input variables as fixed ones. Cosine amplitude scheme is employed to implement the sensitivity analysis as follow (Momeni et al., 2014),

$$R_{ij} = \frac{\sum_{k=1}^{N} (I_{ik} O_{jk})}{\sqrt{\sum_{k=1}^{N} I_{ik}^2 \sum_{k=1}^{N} O_{jk}^2}} \tag{7}$$

where I_i and O_j are the input and output variables, and N is the total number of datasets. R_{ij} value ([0,1]) clarifies the strength of relationship between every input variable and the output in Equation (6). Fig. 4 reveals the results of R_{ij} that as can be seen, T_{am} is the utmost sensitive and significant variable in Equation (6) for forecasting *GWL*.

Albeit it is principally expected that P_m and W_{dm} to be as most influential parameters in *GWL*, according to Table 2 and Fig. 4, T_{am} is qualified as the dominant factor. An explanation for this unconventionality is that, whereas Urmia plain is largely located in a semi-arid part of Iran and consequently perceptible reduction of rainfalls rate and main rivers discharge in this region, T_{am} plays a substantial role in *GWL* and has also significant effect in hydrological studies.

Table 2
Computed weight value for used time series variables with respect to *GWL* by Shannon entropy scheme.

Variables	ET_m	T_{am}	P_m	W_{dm}
Weight Value (%)	2.79	43.47	20.71	33.03

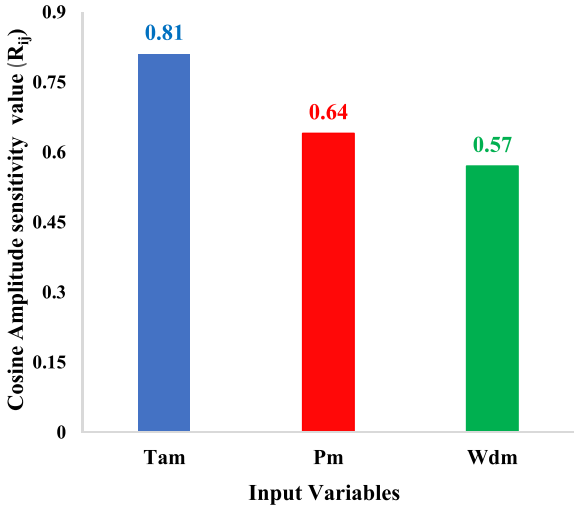


Fig. 4. Results of R_{ij} with respect to input variables.

2.3. Models design

In relation to Table 1 and Fig. 3, it is intelligible that *GWL* has unstable behavior by the kurtosis and skewness. It can be distinguished that there is a high degree of basic complexity nature and non-linearity relationships among the time series variables used in Equation (6). Consequently, robust accurate techniques are required.

As above-mentioned, in this investigation, to forecast the time series *GWL*, three different layer structures of GRU-based neural network models are developed. Since the severe vacillations of *GWL* bring about inaccuracy and adversative effects on the learning capability of models, all datasets initially should be normalized to unit variance and zero mean. Then, all inputs and output variables are separated into two subsections. In this sense, 70 % of total datasets (the first 120 months) are utilized for calibration of models and the remaining (60 months) are applied for validation.

2.3.1. Gated recurrent Unit (GRU) neural network

Typical neural networks presume the input/output variables are autonomous from each other and do not take into consideration the influence of prior information on the inputs that are not appropriate supposition for long-term time series prediction. Nevertheless, RNNs (recurrent neural networks) can remind the prior information and use it for estimating the target. At large, since RNNs can maintain the previous information, they can tentatively handle information in randomly long sequences. But, the ordinary RNNs with backpropagation (BP) training algorithm in the learning long-term dependencies presented practically weak performance and come upon problems in modeling (Athiwaratkun and Stokes, 2017).

GRU employs a particular gating mechanism of recurrent neural networks (RNNs) in which for resolving integral deficits and improving

capability in learning long-period dependencies of the primary RNNs (i.e., vanishing gradients), a specific structure was subjoined to its cell. Fig. 5 explains the functional mechanism and interior memory cell of GRU. In Fig. 5, it is supposed that r_t is reset gate of GRU at time t, x_t is the input at time t, and h_{t-1} is the hidden state at time $t-1$. Moreover, W_r and U_r signify a weight matrix for the input data and hidden state, respectively (Cho et al. 2014). Contrary to LSTM, GRU does not contain isolated memory cells, vice versa it employs a single hidden state ht to distribute info over time steps. As well, the input and forget gates are united with an update gate (z), and r_t is straightly applied to ht_{-1} to attain ht (the candidate state) which is added to ht (Chen et al., 2020). The paradigm of GRU is portrayed via the following equations (Cho et al. 2014):

$$z_t=\sigma(W_z x_t+U_z h_{t-1}+b_z) \tag{8}$$

$$r_t=\sigma(W_r x_t+U_r h_{t-1}+b_r) \tag{9}$$

$$\widetilde{h}_t = \tanh\left(W\widetilde{h}x_t + U\widetilde{h}\,(r_t \times h_{t-1}) + b\widetilde{h}\right) \tag{10}$$

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times \widetilde{h}_t \tag{11}$$

In the above-mentioned equations σ and tanh are the logistic sigmoid and hyperbolic tangent functions, respectively, the symbol " × " characterizes the element-wise multiplication, and b is the bias vector. They are all learnable and unshared factor sets. On account of the sigmoid function impact, all gates are a vector with amounts limited within the range (0, 1). At what time that rt is closed, GRU will overlook ht-1 and are affected just by xt. Also, zt regulates how much information of ht_{-1} can be exceeded to h_t (Cho et al. 2014).

2.3.2. Model development

In this investigation, to progress the ability of designed models, the algorithm tuning technique is operated. For sake of this aim, by tuning the type of *SAF*, *P-rate* value, *NoN*, and *NAD* as main hyperparameters is tried to achieve a suitable configuration to thwart the overfitting problem in the designed models. Since there is not an efficient principle to predetermine appropriate main hyperparameters for a given model on definite data, the current process considers a time-consuming and demanding task. In this direction, numerous scenarios are characterized and the apt main hyperparameters are achieved by the user via a trial-and-error process.

First of all, as suggested by Maier et al. (2010) the framework of ANN models ought to be designed consistent with the intentions of research. In this respect, we initially developed the general Single-GRU layer network model (1) by MATLAB 2021a as is shown in Fig. 6 (a). Then, for the sake of tuning *NAD* hyperparameter, two different deeper structural models (2) and (3), namely GRU2 and GRU2 + models, are developed, as are shown in Fig. 6 (b) and (c). As seen in Fig. 6, the number of layers in the structural models (1), (2), and (3) are 5, 6, and 7, respectively. Indeed, the model (2) is a simple Double-GRU model, whereas the new proposed model (3) is Double-GRU coupled with Addition layer (+) with the module of sequence-to-sequence regression. The model (1) is reflected as our baseline and is applied to compare with the performance of other designed models.

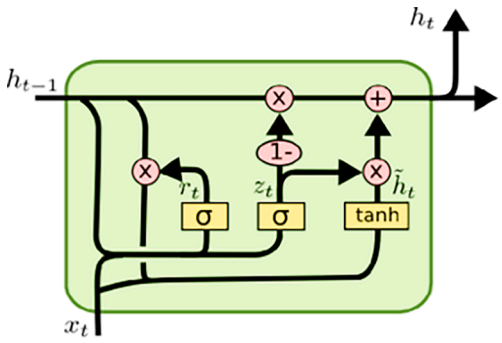


Fig. 5. Internal operations of a GRU cell.

In the structure of designed models, the first layer intakes sequence and time series input variables into the layers' network models. In this respect, T_{am}, P_m, W_{dm} are utilized with time steps standing for months as input variables, so its size in all developed GRU based structures is set to 3. Addition layer adds input variables from several neural network layers' element-wise with an equal dimension. The number of inputs in this layer is regulated indiscreetly by MATLAB software, yet it has merely a sole output (The MathWorks, Inc., 2021a).

DNNs are suitable to analyze large datasets, but too high ones can simply cause overfitting. In connection with this, because Dropout layer offers a functioning regularization system, it was utilized to preclude this problem (Hinton et al., 2012). The topmost influence of Dropout layer is that at each training repetition by updating the system, it discounts systematically or dropout some neurons in operated layers with a probability rate of *P*-value (Srivastava et al., 2014). In the present prediction, to tune *P-rate*, values of 0.4, 0.5, and 0.6 are utilized.

Albeit GRU neural network layer has enough skill to learn long-term time series data, its fitting facility could be inadequate (Zhang et al., 2018), so Fully Connected layer (FC) must be added to the utilized structures. It multiplies the inputs by a weight matrix and next adds a bias vector to expand the fitting ability of designed models (The MathWorks, Inc., 2021a). Regarding, its input size in developed models in the training stage is set to "auto" to specify automatically its size by MATLAB software, and its output size is set to 1.

The ending layer of designed GRU based models is chosen as a regression output layer by which estimates "*half-mean-squared-error loss*" for regression missions by loss function as:

$$Loss = \sum_{i=1}^{N} (y_o - y_p)^2 \tag{12}$$

where y_p and y_o are the predicted and observed value of *GWL* at time *i*, respectively.

To tune *SAF*, various mixtures of *tanh* and *softsign* as the activation function are used in GRU layers and to tune *NoN*, diverse values in each model are tested with the gate activation function of sigmoid.

To update the network bias and weights of the training options in the designed models, "*Adam*" optimization algorithm with 1000 maximum reiterations with an "*Initial Learning Rate*" of 0.5 is selected. Also, to lessen the impact of padding, the value of "*Mini Batch Size*" is set to 20 and "*Gradient threshold*" is also set to 1 to deter the gradients exploding predicament. Indeed, these appointed factors are the norm of the developed models which have a noticeable influence on their performance.

3. Performance evaluation metrics

In this investigation, the coefficient of determination (R^2) and root mean square error (*RMSE*) are utilized to evaluate the accurateness and effectiveness of the designed models in forecasting the time series *GWL*. These statistical metrics are formulated by the following equations as follow:

$$R^2 = \left(\frac{\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)}{N\sigma_x\sigma_y}\right)^2 \tag{13}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(x_i - y_i)^2}{N}} \tag{14}$$

where *N* denotes the number of datasets, y_i and x_i are the forecasted and measured *GWL* at the time *i*, σ_y and σ_x are the standard deviation of the

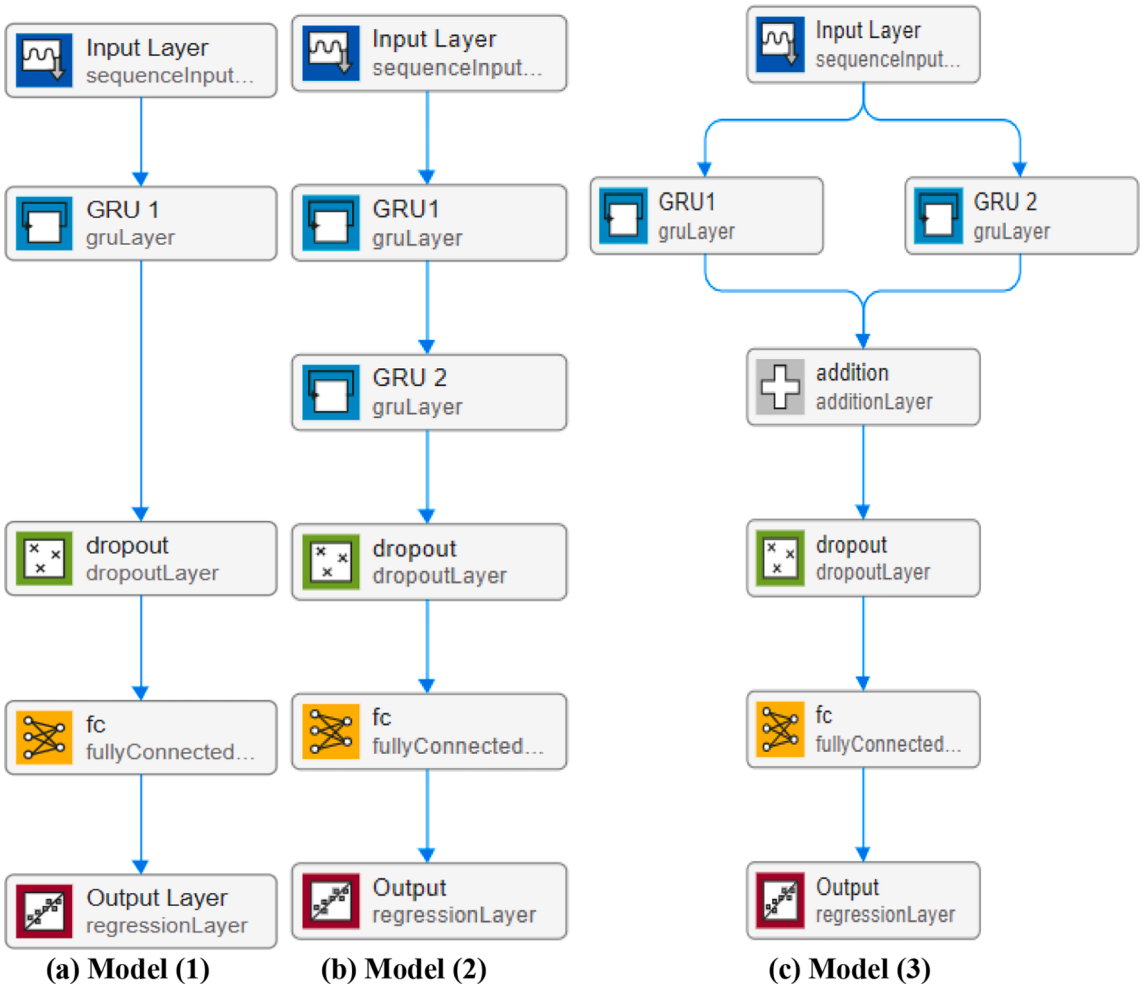


Fig. 6. Architect of GRU based layers network models. (a) Structure of the general Single-GRU layer network model, (b) Structure of simple Double-GRU layers network model (GRU2 model), (c) Structure of Double-GRU model coupled with Addition layer (GRU2+ model).

forecasted and measured *GWL*, respectively, μ_y and μ_x are the forecasted and measured *GWL*, respectively. The optimum amounts for the aforesaid metrics are 1 and 0, respectively.

Table 3
Results of model (1) in the training and testing stages.

Training stage				*Testing stage*			
NoN	*P-rate*	*Running time (s)*		*RMSE(m)*		R^2	
		tanh	*softsign*	*tanh*	*softsign*	*tanh*	*softsign*
50	0.4	36	32	0.2168	0.2201	0.669	0.615
50	0.5	33	27	0.2249	0.2334	0.629	0.593
50	0.6	31	22	0.2224	0.2324	0.647	0.608
75	**0.4**	**49**	**35**	**0.1823**	**0.1925**	**0.796**	**0.669**
75	0.5	41	33	0.2032	0.2106	0.691	0.632
75	0.6	37	30	0.1996	0.2089	0.715	0.651
100	0.4	86	79	0.2109	0.2215	0.563	0.555
100	0.5	78	72	0.2202	0.2322	0.452	0.422
100	0.6	71	68	0.2312	0.2436	0.422	0.405
150	0.4	102	93	0.2814	0.2905	0.354	0.311
150	0.5	94	86	0.3125	0.3212	0.302	0.291
150	0.6	86	81	0.3225	0.3336	0.292	0.272

4. Results and discussion

4.1. Validation of general Single-GRU model

In the model (1) after several experiments, the optimal value for *P-rate*, *SAF*, and *NoN* is attained 0.4, *tanh*, and 75, respectively. Under the optimal scenario, in the training stage after 1000 iterations, the value of *RMSE* and *loss* as the performance metrics are approximately obtained 1E-5 and 1E-7, respectively, however, the value of *learning rate* is obtained 5E-8. After calibration, this model is validated by the testing datasets. Results of model (1) with respect to *P-rate*, *SAF*, and *NoN* in the calibration and validation stages are presented in Table 3 wherein the bolded amounts denote the optimal scenario features.

In keeping with Table 3, it can be understood that too many *NoN* (150) induces to increase *RMSE* because of the overfitting problem, but smaller *NoN* (50) reduces the network learning ability or the underfitting problem. Also, the model (1) in the training stage is more precise than the testing stage.

As can be seen from Table 3, in the same *NoN*, by increasing *P-rate,* the value of running time decreases, yet in the same *NoN* and *P-rate*, the scenarios with *SAF* of *softsign* are faster than *tanh*. Besides, in the same *P-rate,* by increasing the value of *NoN*, the running time increases.

4.2. Validation of Double-GRU models

In the training stage of Double-GRU models, under the optimal scenario after 1000 iterations, the value of *RMSE* and *loss* as the performance metrics for simple Double-GRU model (2) (GRU2 model) are obtained about 1E-4 and 1E-5, respectively, and for the new proposed model (3) about 1E-5 and 1E-7, respectively. Besides, the value of *learning rate* for both models is obtained 5E-8.

As abovementioned, different combinations of activation functions were used in *SAF* of GRU layers in the hidden layer of both models. In this respect, in both models after several trials, the "*tanh-softsign*" combinations are specified as an apt option. Also, the ideal value of *P-rate* and *NoN* for both models are achieved 0.5 and 50, respectively.

Results of models (2) and (3) with respect to *P-rate* and *NoN* are presented in Table 4 wherein the expression "*tanh-softsign*" implies that the kind of *SAF* in GRU layers 1 and 2 are the *tanh* and *softsign*, respectively, and also the bolded quantities denote to the optimal scenarios.

According to Table 4, it is also clear that in both models too many *NoN* (150) causes to increase *RMSE* because of the overfitting problem, nevertheless, the smaller *NoN* (30) causes the underfitting problem. As well, both models in the calibration phase are more precise than the validation phase. Also, in the same *NoN*, by rising *P-rate*, the value of running time decreases, however in the same *P-rate*, by increasing the value of *NoN*, the running time increases.

Over and above, as said in Table 4, it can be discerned that by using dissimilar *AF* in the hidden layers of employed models, they can learn more complex and nonlinear functions. With regards to, the "*tanh-softsign*" combination causes the model not to be more susceptible to the overfitting problem in the training phase. *Softsign* activation function can strengthen the rapidity of models training, while *tanh* activation function can suitably distinguish intricate relationships in the long-term time series (Yin et al., 2020). However, in this case, the kind of data has a momentous impact in figuring which combination ponders as the best one.

4.3. Performance evaluation of designed models

In the current research, we emphasize the capacity and infrastructure of designed models to assess their performance. Relating to, MATLAB 2021a offers a significant parameter namely, *TLP* (Total Learnable Parameters), as a main measure and norm to determine the physical capacity of designed models. A model with optimum *NAD* and *NoN* definitely results in a good-adjusted *TLP* and consequently reduces the impacts of overfitting problem. Concerning, Table 5 is provided to

Table 5
Characteristics and statistical indices in the optimal scenario of applied models in the testing stage.

Model	*NoN*	*RMSE*(m)	R^2	*Running Time* (s)	*TLP*
1	75	0.1823	0.796	49	17851
2	50	0.279	0.56	53	23301
3	**50**	**0.094**	**0.91**	**47**	**16251**

Table 4
Results of models (2 and 3) in the training and testing stages.

Training stage					*Testing stage*			
NoN	*P-rate*	SAF_1-SAF_2	*Running Time* (s)		*RMSE* (m)		R^2	
			Model (2)	Model (3)	Model (2)	Model (3)	Model (2)	Model (3)
30	0.4	*tanh-softsign*	56	53	0.2998	0.1417	0.48	0.76
30	0.5	*tanh-softsign*	51	47	0.2905	0.1391	0.52	0.81
30	0.6	*tanh-softsign*	49	44	0.3012	0.1499	0.43	0.75
50	0.4	*tanh-softsign*	60	57	0.2814	0.1218	0.51	0.86
50	**0.5**	***tanh-softsign***	**53**	**47**	**0.279**	**0.094**	**0.56**	**0.91**
50	0.6	*tanh-softsign*	50	45	0.2915	0.1121	0.52	0.88
75	0.4	*tanh-softsign*	75	69	0.2845	0.1304	0.46	0.82
75	0.5	*tanh-softsign*	68	64	0.2831	0.1258	0.53	0.85
75	0.6	*tanh-softsign*	61	57	0.2918	0.1489	0.51	0.78
100	0.4	*tanh-softsign*	101	94	0.2988	0.1721	0.41	0.68
100	0.5	*tanh-softsign*	94	86	0.2928	0.1613	0.49	0.72
100	0.6	*tanh-softsign*	88	80	0.3012	0.1887	0.44	0.7
150	0.4	*tanh-softsign*	169	153	0.3589	0.3087	0.39	0.55
150	0.5	*tanh-softsign*	158	145	0.3435	0.2914	0.41	0.59
150	0.6	tanh-softsign	147	132	0.3619	0.3105	0.31	0.57

contribute the features of optimal scenario of designed models in the testing stage.

In keeping with Table 5, the model (2) on account of the maximum value of *TLP* is instinctively anticipated to be the most precise and the best one. All the same, in keeping with the evaluation metrics, it can be noticed that in the same hyperparameters the new proposed model (3) outperforms and predicts more exactly than models (2) and (1). The performance statistics of the model (3) has resulted in an R^2 of 0.91 and an *RMSE* of 0.092 (m), whilst the model (2) and (1) has resulted in an R^2 of 0.56 and 0.796, an *RMSE* of 0.279, and 0.1823 (m), respectively. The chief reason for this disagreement and quirkiness can be explained for too extra value of *TLP* in the model (2) compared to the model (3) and (1). Videlicet, the model (2) because of overcapacity memorizes partially the training dataset, meaning it relatively overfits and losses in the optimization process. In sum, the model (2) shows weak performance and is not suitable to estimate the time series *GWL* in the study area.

All in all, according to the value of *TLP* in the model (3), it can be concluded that in the same hyperparameters, inserting Addition layer (+) into the model (2) causes to create a suitable *NAD* with a balanced *TLP* value. Furthermore, it cannot be always concluded that only by increasing the number of GRU layers and *NoN*, the accuracy and capability of designed models are undoubtedly improved. Regarding, as a topmost result, for achieving an operative model, an appropriate structural architecture with optimal *NoN* and well-proportioned *TLP* along with utilizing the algorithm tuning process must be employed.

In terms of iterations and running time in the designed models, the model (3) attributable to the balanced *TLP* value begets the model to acquire quicker sub-optimum weight sets (1000 iterations in the 47 s). Nonetheless, the GRU2 model with the highest *TLP* induces the model to acquire further and even generally optimal weight sets and consequently takes longer to train (1000 iterations in the 53 s).

Fig. 7 demonstrates the schematical comparison of the observed and predicted time series *GWL* by the designed models in the testing stage. Visual analysis of Fig. 7 affirms that the model (3) can suitably remember the previous observed time series *GWL* and fittingly captures the variation trend. However, the GRU2 model fails to fit data and diverges from the observed values, particularly in the peak values that indicate to the high deviations.

For determining the superior model, it is necessary to consider the effect of running time scale. In this direction, the "Model Grading" process proposed by Vaheddoost et al. (2016) is employed as the supplementary statistical comparison for developed models. In this process, success grade (*SG*) and failure grade (*FG*) of performance is defined as follows:

Willmott index (*WI*)

$$WI = 1 - \frac{\sum_{i=1}^{N}(x_i - y_i)^2}{\sum_{i=1}^{N}(|y_i - \mu_x| + |x_i - \mu_x|)^2} \tag{15}$$

Nash-Sutcliffe efficiency (*NSE*)

$$NSE = 1 - \frac{\sum_{i=1}^{N}(x_i - y_i)^2}{\sum_{i=1}^{N}(x_i - \mu_x)^2} \tag{16}$$

$$SG(WI) = \frac{WI_i}{WI_{\max}} \times 10 \tag{17}$$

$$SG(NSE) = \frac{NSE_i}{NSE_{\max}} \times 10 \tag{18}$$

$$FG(RMSE) = \frac{RMSE_i}{RMSE_{\max}} \times 10 \tag{19}$$

$$FG(Time) = \frac{Time_i}{Time_{\max}} \times 10 \tag{20}$$

Total Grade (*TG*) of each model obtains by adding up *FG* and *SG* of each model separately which can change between −20 and + 20, and is defined as follow,

$$TG = SG(WI) + SG(NSE) - FG(RMSE) - FG(Time) \tag{21}$$

TG obtained by Equations 15–21 is presented in Table 6. Along with the values of *TG* in Table 6, the model (3), is selected as the superior model for forecasting the time series *GWL*, whereas the results gained by the model (1) are the second best. It shows that the newly suggested model (3) outperforms and regards as the best and most accurate model.

5. Conclusion

Knowledge of predicting groundwater level fluctuations is essential for operative managing of drinking water exclusively in arid and semi-arid regions, surface subsiding, and security of water resources. In this sense, in the current research, we evaluated *GWL* fluctuations over a long-term in Urmia plain a farming region in Urmia city, West Azerbaijan Province, Northwest Iran, by using three different layer structures of GRU based neural network models via the seq2seq module, modern deep learning scheme. To develop models for precise prediction of *GWL* with time series features, 180 mean monthly observed water table depths of 42 different piezometers during the period (Oct

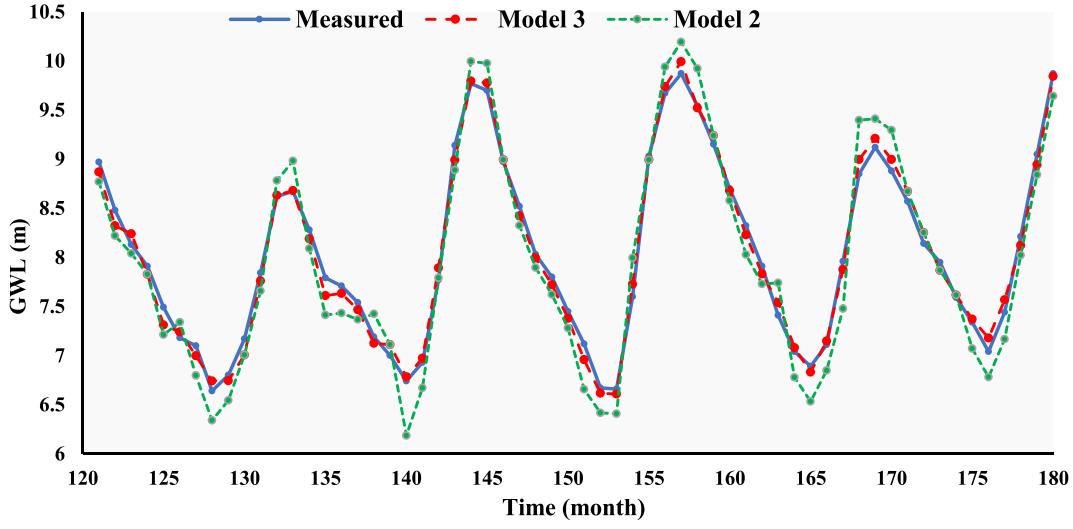


Fig. 7. Comparison of predicted and measured *GWL* in the testing stage (60 months between Oct 2012–Sep 2017).

Table 6
Comparison of optimal developed models in the testing stage and related grades.

Model	*WI*	*NSE*	*Running Time* (s)	*RMSE*	*SG* (SG(WI)+SG(NSE))	*FG* (FG(RMSE)+FG(Time))	*TG*
1	0.9	0.87	49	0.1823	19.02	-15.78	3.25
2	0.85	0.8	53	0.279	17.83	-20	-2.16
3	**0.94**	**0.91**	**47**	**0.094**	**20**	**-12.24**	**7.76**

2002–Sep 2017) were used. The situation of the applied piezometers was selected so that cover all study regions with different statistical indices. For implementing designed models, regional average monthly air temperature (T_{am}), precipitation (P_m), evaporation (ET_m), and total (sum) mean monthly water diversion discharge (W_{dm}) of main rivers collected from different meteorological and hydrometry stations were evaluated as potential variables. However, results of data preprocessing by Shannon entropy scheme confirmed P_m, T_{am}, and W_{dm} as the effective input variables. Yet, T_{am} was nominated as the most influential and sensitive one by Cosine amplitude sensitivity analysis. For better configuration and decreasing overfitting or underfitting problems, algorithm tuning by *SAF*, *P-rate*, and *NAD* as hyperparameters accompanied by the trial-and-error process was adopted. The major outcomes of modelling are as follows:

1. All developed models in the calibration stage were more precise than the validation stage. Also, in the same *NoN*, by rising *P-rate*, the value of running time reduced, yet in the same *P-rate*, by increasing the value of *NoN*, the running time increased.
2. In Double-GRU models after several experiments, "*tanh-softsign*" combinations in *SAF* of GRU layers in the hidden layer were specified as an apt option. Also, the optimal value for *P-rate* and *NoN* in both models were achieved as 0.5 and 50, respectively.
3. Despite the maximum value of *TLP* in the model (2), the new proposed model (3) excelled in the models (2) and (1). The models (3) in comparison with the models (2) and (1) makes to diminish the value of *RMSE* almost by %66 and %48, respectively.
4. Because of the poor performance of the model (2), as a substantial outcome, it can be reasoned that for accomplishing an effective model, the optimum *NAD*, *NoN*, and *TLP* through the algorithm tuning and trial-and-error processes must be utilized.

The recommended new operational GRU-based structure (3) by the authors considers exceptional from usual neural network models and has not been regularly employed in the field of hydrology science. It presents a satisfactory state-of-the-art approach since its superiority was proven by practical performance. Counter to the other AI algorithms, it considers so inexpensive, uncomplicated, and time-saving. As well, in the same hyperparameters, owing to the well-proportioned *NAD*, it caused a balanced *TLP* value and consequently more precision in comparison with the traditional GRU-based models, yet it was taken too long to train and converge compared to the model (2) even with less *NoN*. Accordingly, it can be employed as a smart intellectual model to estimate time series *GWL* under diverse conditions of water resources management.

This model under the optimal hyperparameters yielded in an R^2 of 0.91, an *RMSE* of 0.094 (m), and an *NSE* of 0.91. However, in the developed models by Wunsch et al. (2021) under the best method i.e., NARX via the seq2seq module and meteorological input parameters, GWL modeling in the applied wells dependence on the range of groundwater resulted in an R^2 of among 0.08 and 0.72, an *RMSE* of among 0.07 and 0.52 (m), and an *NSE* of among −0.07 and 0.9.

In spite of the remarkable advantages of the novel model (3), it has some limits such as prerequisite to the continual gauge of different water table depths, a high number of monitoring piezometers, hydrological, and meteorological record of data over a long-time period to estimate accurately *GWL* in a study area. Also, because *GWL* cannot deliberate as a stationary typical hydrological time series or a consistent natural phenomenon, forecasting it with different interval responses in an agricultural area is definitely regarded as a demanding and time-consuming problem.

Notwithstanding the present investigation appraised the capability of different structures of GRU-based model in estimating *GWL*, the approaching studies could be performed by other types of methods via hybridizing DNN models with optimization algorithms such as Particle Swarm Optimization (PSO), Fertile Field Algorithm (FFA), etc. Then, the consequences should be paralleled with the upshots of the present study so as to the excellent AI technique could be specified.

To end, consistent with the consequences of the current modelling, the authors advise *TLP* factor as a noteworthy criterion and critical measure to assess the forecasting capability and performance of designed DNNs-based models. Indeed, it determines the practical capacity and can be applied to check the overfitting or underfitting problems.

Funding

This study was funded by "Hasan Kalyoncu University" and "Shahid Chamran University of Ahvaz".

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

We are grateful to the Hasan Kalyoncu University and Research Council of Shahid Chamran University of Ahvaz for financial support (GN: SCU.WI98.280). Also, we are also grateful to Hamid Saadatnejad gharahassanlou for association.

References

Alabjah, B., Amraoui, F., Chibout, M., Slimani, M., 2018. Assessment of saltwater contamination extent in the coastal s of Chaouia (Morocco) using the electric recognition. J. Hydrol. 566, 363–376. https://doi.org/10.1016/j.jhydrol.2018.09.003.

Athiwaratkun, B., Stokes, J.W., 2017, March. Malware classification with LSTM and GRU language models and a character-level CNN. In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 2482-2486). IEEE.

Bowes, B.D., Sadler, J.M., Morsy, M.M., Behl, M., Goodall, J.L., 2019. Forecasting groundwater level in a flood prone coastal city with long short-term memory and recurrent neural networks. Water 11 (5), 1098.

Chang, F.J., Chang, L.-C., Huang, C.W., Kao, I.F., 2016. Prediction of monthly regional groundwater levels through hybrid soft-computing techniques. J. Hydrol. 2016 (541), 965–976.

Chen, M., Izady, A., Abdalla, O.A., 2017. An efficient surrogate-based simulation-optimization method for calibrating a regional MODFLOW model. J. Hydrol. 544, 591–603.

Chen, Y., Liu, G., Huang, X., Chen, K., Hou, J., Zhou, J., 2020. Development of a surrogate method of groundwater modeling using gated recurrent unit to improve the efficiency of parameter auto-calibration and global sensitivity analysis. J. Hydrol., 125726

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

Coulibaly, P., Anctil, F., Aravena, R., Bobée, B., 2001. Artificial neural network modeling of water table depth fluctuations. Water Resour. Res. 37 (4), 885–896.

Daliakopoulos, I.N., Coulibaly, P., Tsanis, I.K., 2005. Groundwater level forecasting using artificial neural networks. J. Hydrol. 309 (1), 229–240.

Dalkiliç, H.Y., Gharehbaghi, A., 2021. numerical modeling of groundwater flow based on explicit and fully implicit schemes of finite volume method. Journal of Engineering Research 9 (4B), 56–69. https://doi.org/10.36909/jer.9253.

EARWO (East Azerbaijan Regional Water Organization). 2020. Preparation of water balance and water cycle in the Malekan region.56p.

Feng, S., Kang, S., Huo, Z., Chen, S., Mao, X., 2008. Neural networks to simulate regional groundwater levels affected by human activities. Ground Water 46 (1), 80–90.

Gao, S., Huang, Y., Zhang, S., Han, J., Wang, G., Zhang, M., Lin, Q., 2020. Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. J. Hydrol. 589, 125188.

Gharehbaghi, A., 2016. Explicit and implicit forms of differential quadrature method for advection– diffusion equation with variable coefficients in semi-infinite domain. J. Hydrol. 541(B), 935–940.

Gharehbaghi, A., 2017. Third- and fifth-order finite volume schemes for advection–diffusion equation with variable coefficients in semi-infinite domain. Water and Environment Journal.

Gharehbaghi, A., 2022. Fully implicit form of differential quadrature method for multi-species solute transport in porous media, Teknik Dergi, 33(4), 10.18400/tekderg.975457 (in press).

Ghasemlounia, R., Gharehbaghi, A., Ahmadi, F., Saadatnejadgharahassanlou, H., 2021. Developing a novel framework for forecasting groundwater level fluctuations using Bi-directional Long Short-Term Memory (BiLSTM) deep neural network. Computers and Electronics in Agriculture 191, 106568.

Graves, A., 2012. Supervised Sequence Labelling with Recurrent Neural Networks. Springer-Verlag, Berlin, Heidelberg.

Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R., 2012. Improving neural networks by preventing co-adaptation of feature detectors. Comput. Sci. 3, 212–223.

Houéménou, H., Tweed, S., Dobigny, G., Mama, D., Alassane, A., Silmer, R., Babic, M., Ruy, S., Chaigneau, A., Gauthier, P., Socohou, A., 2020. Degradation of groundwater quality in expanding cities in West Africa. A case study of the unregulated shallow aquifer in Cotonou. J. Hydrol. 582, 124438.

Jeong, J., Park, E., 2019. Comparative applications of data-driven models representing water fluctuations. J. Hydro. 572, 261–273.

Jeong, J., Park, E., Chen, H., Kim, K.Y., Han, W.S., Suk, H., 2020. Estimation of groundwater level based on the robust training of recurrent neural networks using corrupted data. J. Hydrol. 582, 124512.

Liang, C., Hongqing, L., Mingjun, L., Qingyun, D.u., 2018. Dongting Lake water level forecast and its relationship with the Three Gorges Dam based on a long short-term memory network. Water 10 (10), 1389.

Maier, H.R., Jain, A., Dandy, G.C., Sudheer, K.P., 2010. Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions. Environ. Model. Softw. 25 (8), 891–909.

Meshgi, A., Schmitter, P., Babovic, V., Chui, T.F.M., 2014. An empirical method for approximating stream baseflow time-series using groundwater fluctuations. J. Hydrol. 519, 1031–1041. https://doi.org/10.1016/j.jhydrol.2014.08.033.

Mirzavand, M., Khoshnevisan, B., Shamshirband, S., Kisi, O., Ahmad, R., Akib, S. 2015. Evaluating groundwater level fluctuation by support vector regression and neuro-fuzzy methods: a comparative study. Natural Hazards 1, no. 1 (2015): 1-15.

Mohanty, S., Jha, M.K., Kumar, A., Panda, D.K., 2013. Comparative evaluation of numerical model and artificial neural network for simulating groundwater flow in Kathajodi-Surua Inter-basin of Odisha, India. J. Hydrol. 2013 (495), 38–51.

Momeni, E., Nazir, R., Jahed Armaghani, D., Maizir, H., 2014. Prediction of pile bearing capacity using a hybrid genetic algorithm-based ANN. Measurement 57, 122–131.

Pan, M., Zhou, H., Cao, J., Liu, Y., Hao, J., Li, S., Chen, C.H., 2020. Water Level Prediction Model Based on GRU and CNN. IEEE Access 8, 60090–60100.

Park, C., Lee, C., Hong, L., Hwang, Y., Yoo, T., Jang, J., Hong, Y., Bae, K.H., Kim, H.K., 2019. S2-Net: machine reading comprehension with SRU-based self-matching networks. ETRI J. 41, 371–382. https://doi.org/10.4218/etrij.2017-0279.

Raghavendra, N.S., Deka, P.C. 2016. Multistep ahead groundwater level time-series forecasting using Gaussian process regression and ANFIS. Advanced Computing and Systems for Security, Volume 396 of the Series Advances in Intelligent Systems and Computing, pp. 289–302.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by backpropagating errors. Nature 323, 533–536.

Sahoo, S., Jha, M.K., 2013. Groundwater-level prediction using multiple linear regression and artificial neural network techniques: a comparative assessment. Hydrogeol. J. 2013 (21), 1865–1887.

Shannon, C.E., 1948. A mathematical theory of communication. The Bell System Technical Journal 27 (3), 379–423.

Shen, C., 2018. A transdisciplinary review of deep learning research and its relevance for water resources scientists. Water Resour. Res. 54 (11), 8558–8593.

Shiri, J., Kisi, O., Yoon, H., Lee, K.K., Hossein Nazemi, A., 2013. Predicting groundwater level fluctuations with meteorological effect implications—a comparative study among soft computing techniques. Comput. Geosci. 56, 32–44.

Solomatine, D.P., Obstfeld, A., 2008. Data-driven modelling: Some past experiences and new approaches. Hydroinform. 2008 (10), 3–22.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn Res. 15, 1929–1958.

MATLAB User's Guide 2021a, The MathWorks Inc. (Deep Learning Toolbox). Natick, Massachusetts, United State; (2021). Computer Software. www.mathworks.com/.

Vaheddoost, B., Aksoy, H., Abghari, H., 2016. Prediction of water level using monthly lagged data in Lake Urmia, Iran. Water Resour. Manage. 30 (13), 4951–4967.

Vu, M.T., Jardani, A., Massei, N., Fournier, M., 2020. Reconstruction of missing groundwater level data by using Long Short-Term Memory (LSTM) deep neural network. J. Hydrol., 125776

Wunsch, A., Liesch, T., Broda, S., 2021. Groundwater level forecasting with artificial neural networks: a comparison of long short-term memory (LSTM), convolutional neural networks (CNNs), and non-linear autoregressive networks with exogenous input (NARX). Hydrol. Earth System Sci. 25 (3), 1671–1687.

Yin, J., Deng, Z., Ines, A.V., Wu, J., Rasu, E., 2020. Forecast of short-term daily reference evapotranspiration under limited meteorological variables using a hybrid bi-directional long short-term memory model (Bi-LSTM). Agric. Water Manage. 242, 106386.

Yoon, H., Jun, S.C., Hyun, Y., Bae, G.O., Lee, K.K., 2011. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal. J. Hydrol. 396 (1–2), 128–138.

Zhang, J., Zhu, Y., Zhang, X., Ye, M., Yang, J., 2018. Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. J. Hydrol. 561, 918–929.